

TAUHID KHAN

📍 Mumbai, India 📞 +91 8268637307

✉ mail2tauhidkhan@gmail.com

🐙 [GitHub](#) [LinkedIn](#)

🎓 [Google Scholar](#)

SUMMARY

ML Research Engineer with 4+ years of experience developing production computer vision, generative models, and efficient inference systems. Combines this background with hands-on robot learning through mini-pi0, including Flow Matching policies in simulation and autonomous visuomotor policy deployment on a self-assembled SO-101 arm. Experience spans teleoperated demonstration collection, policy training, and physical manipulation experiments, supported by strong expertise in distributed PyTorch training, model distillation, and TensorRT optimization.

RESEARCH INTERESTS

Robot Learning, Robotic Manipulation, Vision-Language-Action Models, Visuomotor Policy Learning, Imitation Learning, Reinforcement Learning, Generative Action Policies, Flow Matching, Multimodal Learning, Computer Vision, Representation Learning, Temporal Action Modeling, Efficient Robot Inference, Simulation-Based Training, Sim2Real Transfer

PROFESSIONAL EXPERIENCE

ML Research Engineer

Sept 2023 – Present

Fynd ShopSense

- Developed and deployed an efficient **Flow Matching-based super-resolution model** conditioned on low-resolution (LR) inputs, combining **local patch-level features with whole-image semantic context** to recover high-frequency details while preserving global structure across diverse image domains.
- Curated a high-resolution training dataset spanning **documents, human faces, landscapes, graphic designs, artwork, and general imagery**, building preprocessing, degradation, filtering, and data-quality pipelines for robust super-resolution training.
- Distilled the multi-step Flow Matching super-resolution model into a **single-step generator** by learning the underlying flow maps, substantially reducing sampling complexity and enabling efficient production inference.
- Designed a memory-efficient tiled inference pipeline for **4× super-resolution**, processing a **2500×2500** low-resolution input using **512×512 patches with 32-pixel overlap** in approximately **4–5 seconds on a single NVIDIA L4 24GB GPU**.
- Converted the complete super-resolution inference stack to **TensorRT**, optimizing all major model components and achieving approximately **40–45% lower inference latency** while maintaining the deployed model's output quality.
- Developed and fine-tuned large-scale diffusion models (**SDXL, Flux.Dev**) using distributed multi-GPU PyTorch training, improving controllable image generation quality by **20%+** based on internal evaluation metrics.
- Architected spatio-temporal video segmentation and inpainting pipelines for foreground object removal and video restoration, enabling processing of videos up to **4K resolution** with stable temporal consistency.
- Led optimization and deployment of a foreground removal pipeline by converting segmentation and inpainting models to **TensorRT** with custom GPU plugins, improving end-to-end inference latency by up to **56%** and reducing compute cost by up to **52%** across resolutions from 720p to 4K.
- Developed production-grade AI media services for image and video generation, restoration, super-resolution, and foreground removal, serving approximately **20K daily active users**.
- Engineered controllable image generation pipelines using IP-Adapters and Textual Inversion across **25+ artistic domains**, expanding model versatility and controllable style generation.

Computer Vision Engineer - I

Feb 2022 – Sept 2023

Wobot Intelligence Pvt. Ltd.

- Reduced production inference CPU utilization from **90% to 40%** by redesigning multi-threaded video processing pipelines, enabling more efficient resource utilization and scalable deployments.
- Developed and deployed multi-instance, multi-camera person and vehicle tracking pipelines using object detectors from the YOLO family and DeepSORT/ByteTrack-based trackers for drive-thru journey-time analytics, serving **1000+ cameras across 250+ locations** with robust occlusion-aware tracking.
- Developed a reusable internal computer vision deployment framework and productionized GPU inference services using Docker and **NVIDIA Triton Inference Server**.

EDUCATION

Bachelor of Science in Computer Science

2022

Mumbai University

CGPA: 9.21/10

Relevant Coursework: Artificial Intelligence, Optimization Techniques, Digital Image Processing, Data Science

PUBLICATIONS

Isokinetic Flow Matching for Pathwise Straightening of Generative Flows

2026

Tauhid Khan

Accepted at the **ICML 2026 Workshop on Structured Probabilistic Inference & Generative Modeling (SPIGM)**

📄 [arXiv:2604.04491](#)

PROJECTS

mini-pi0 – Visuomotor Learning and SO-101 Manipulation

2026

PyTorch, Flow Matching, ViT, ManiSkill, LeRobot, SO-101

🔗 [Source Code](#) [Autonomous Demo](#)

- Developed image-conditioned Flow Matching action-chunk policies, achieving **95.5% success over 200 StackCube-v1 simulation episodes** in ManiSkill with a Transformer + ViT configuration.
- Assembled and calibrated an **SO-101 arm**; implemented gamepad teleoperation with inverse kinematics, bounded joint commands, and dual-camera demonstration recording using LeRobot.
- Trained and deployed a **32M-parameter ViT visuomotor policy** on the SO-101, demonstrating **autonomous manipulation on physical hardware** from collected demonstrations; published videos and the teleoperation and policy-inference code.

TinyML Audio Classification on Edge Devices (Final Year Project)

2022

TensorFlow Lite Micro, Arduino Nano BLE Sense, Embedded AI

📄 [Report](#) 🔗 [Source Code](#)

- Implemented INT8-quantized DCNN models for edge audio classification operating under **20KB memory constraints**.
- Built a TinyML deployment pipeline for the Arduino Nano BLE Sense as part of the Google TensorFlow Lite Micro Challenge and university final-year project.

TECHNICAL SKILLS

Programming: Python, Linux Shell

Robotics & Robot Learning: Robot Learning, Robotic Manipulation, Vision-Language-Action Models, Visuomotor Policies, Imitation Learning, Reinforcement Learning (PPO, SAC Fundamentals), Flow Matching Policies, ManiSkill, MuJoCo, Isaac Sim, Robosuite, LeRobot

Computer Vision & Multimodal: PyTorch, OpenCV, Vision-Language Models, Multimodal Learning, Representation Learning, Image Segmentation, Video Segmentation, Object Detection, Multi-Object Tracking, Image/Video Restoration
Generative AI: Diffusion Models, Flow Matching, SDXL, Flux.Dev, Super-Resolution, Flow Distillation, One-Step Generative Models, Image-to-Image Generation, IP-Adapters, Textual Inversion, Parameter-Efficient Fine-Tuning (LoRA/PEFT), Image Generation, Video Generation, Controllable Generation, Multimodal Generation

Training & Infrastructure: Distributed Training (DDP, FSDP), TensorFlow, Docker, NVIDIA Triton, TensorRT, FastAPI, Git, MLflow, Weights & Biases

LANGUAGES

English (Professional Proficiency), Hindi (Native)